Application of Biclustering to the Discovery of Constant and Gradual Patterns

UNIVERSITÉ DE LORRAINE

Nyoman Juniarta, Miguel Couceiro, Amedeo Napoli, and Chedy Raïssi

Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

nyoman.juniarta@loria.fr

cnrs

Loria Laboratoire lorrain de recherche en informatique et ses applications

1. Introduction

Biclustering is a **simultaneous grouping** of rows and columns of a matrix. Here we are focusing on **constant-column** and **coherent-sign-changes** biclustering. We show that those types of biclustering can be used in the problem of **gradual pattern mining**.

2. Biclustering

Difference between clustering and biclustering

Clustering

	A	B	C	D	E
1					
2					
3					
4					
5					

Grouping of rows

Biclustering

	A	B	C	D	E
1					
2					
3					
4					
5					

Grouping of rows and columns

Some types of bicluster

Constant-value bicluster

2	2	2	2
2	2	2	2
2	2	2	2
2	2	2	2

Constant-column (CC) bicluster

2	5	1	6
2	5	1	6
2	5	1	6
2	5	1	6

Coherent-sign-changes (CSC) bicluster

+	+	-	+
+	+	-	+
-	-	+	-
+	+	-	+

All rows jointly are in a given "agreement" with the columns

Relation of CC and CSC biclusters

	a	b	c	d	e
1	-	-	-	-	+
2	-	-	-	-	+
3	-	-	-	+	+
4	+	+	-	+	+
5	-	+	+	-	+
6	+	-	+	-	-

CSC bicluster
{4, 5}, {a, c, d}
=
CC bicluster
{4, 5}, {ac, ad}

	ab	ac	ad	ae	bc	bd	be	cd	ce	de
1	S	S	S	D	S	S	D	S	D	D
2	S	S	S	D	S	S	D	S	D	D
3	S	S	D	D	S	D	D	D	D	S
4	S	D	S	S	D	S	S	D	D	S
5	D	D	S	D	S	D	S	D	S	D
6	D	S	D	D	S	D	D	S	S	D

S = same
D = different

3. Partition Pattern Structures for CC Biclustering

	m_1	m_2	m_3	m_4	m_5
g_1	1	5	3	2	7
g_2	1	1	4	2	7
g_3	2	5	4	5	3
g_4	2	5	4	5	7

Partition (of object)

A partition $d = \{p_i\}$ of a set of objects G is a collection of $p_i \subseteq G$ such that there is no overlap among p_i.
Given G as the set of objects and M as the set of attributes, then a partition function $\delta: M \to D$ maps an attribute m to a partition d.
Example:
$\delta(m_1) = \{\{g_1, g_2\}, \{g_3, g_4\}\}$

Similarity and order between two partitions

$d_1 = \{p_i\}$ and $d_2 = \{p_j\}$

$$d_1 \sqcap d_2 = \bigcup_{i,j} \{p_i \cap p_j\}$$

$$d_1 \sqsubseteq d_2 \Leftrightarrow d_1 \sqcap d_2 = d_1$$

Example:
$\delta(m_1) = \{\{g_1, g_2\}, \{g_3, g_4\}\}$
$\delta(m_2) = \{\{g_1, g_3, g_4\}, \{g_2\}\}$
$\delta(m_1) \sqcap \delta(m_2) = \{\{g_1\}, \{g_2\}, \{g_3, g_4\}\}$

Partition pattern concept and CC bicluster

Given a set of attribute M and partition space D, a pair (A, d) is a partition pattern concept iff $A' = d$ and $d' = A$, where:

$A' = \sqcap_{m \in A}\, \delta(m)$ $\qquad A \subseteq M$

$d' = \{m \in M \mid d \sqsubseteq \delta(m)\}$ $\qquad d \in D$

Example:

A concept $(\{m_1, m_4\}, \{\{g_1, g_2\}, \{g_3, g_4\}\})$

Two CC biclusters: $(\{g_1, g_2\}, \{m_1, m_4\})$ and $(\{g_3, g_4\}, \{m_1, m_4\})$

4. Gradual Patterns

Hotel	City population	Distance from center	Price
h_1	1000	5	45
h_2	500	10	25
h_3	600	1	50
h_4	40000	2	100

The hotels in a **big cities** are **more expensive** than those in a **smaller cities**.
The hotels **near city center** are **more expensive** than those **far from center**.

Attribute	h_1h_2	h_1h_3	h_1h_4	h_2h_3	h_2h_4	h_3h_4
Pop.	>	>	<	<	<	<
Dist.	<	>	>	>	>	<
Price	>	<	<	<	<	<

CSC bicluster

Four pairs of hotels (out of six) show the relation among city population, hotel's distance from city center, and hotel's price.

5. References

1. Madeira, S. C., & Oliveira, A. L. (2004). Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, *1*(1), 24-45.
2. Codocedo, V., & Napoli, A. (2014, August). Lattice-based biclustering using Partition Pattern Structures. In *ECAI* (Vol. 14, pp. 213-218).
3. Di-Jorio, L., Laurent, A., & Teisseire, M. (2009, August). Mining frequent gradual itemsets from large databases. In *International Symposium on Intelligent Data Analysis* (pp. 297-308). Springer, Berlin, Heidelberg.